Systems & Control Letters 179 (2023) 105603

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

A "fundamental lemma" for continuous-time systems, with applications to data-driven simulation

P. Rapisarda [a,*], M.K. Çamlibel [b], H.J. van Waarde [b]

[a] School of Electronics and Computer Science, University of Southampton, United Kingdom
[b] Bernoulli Institute for Mathematics, Computer Science, and Artificial Intelligence, University of Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 20 February 2023
Received in revised form 18 July 2023
Accepted 20 July 2023
Available online 7 August 2023

Keywords:
Persistency of excitation
Continuous-time linear time-invariant systems
Chebyshev polynomials
Data-driven simulation

ABSTRACT

We are given one input–output (i-o) trajectory (u, y) produced by a linear, continuous time-invariant system, and we compute its Chebyshev polynomial series representation. We show that if the input trajectory u is sufficiently persistently exciting according to the definition in Rapisarda et al. (2023), then the Chebyshev polynomial series representation of *every* i-o trajectory can be computed from that of (u, y). We apply this result to data-driven simulation of continuous-time systems.

1. Introduction

Before stating the problem we aim to solve and illustrating our contributions, we briefly recall the discrete time "fundamental lemma" of [1] and explain its relevance for data-driven simulation and control of discrete-time systems. Let $\left\{\begin{bmatrix}\widehat{u}_k\\ \widehat{y}_k\end{bmatrix}\right\}_{k=0,\dots,T}$ be input–output (i-o) samples produced by a system $\mathfrak{B}$ represented by a system of linear, constant-coefficient difference equations, with $\widehat{u}_k \in \mathbb{R}^m$ and $\widehat{y}_k \in \mathbb{R}^p$, $k = 0, \dots, T$. Arrange the data in a *Hankel matrix of depth L*:

$$\mathfrak{H}_L(\widehat{u},\widehat{y}) := \begin{bmatrix} \widehat{u}_0 & \widehat{u}_1 & \dots & \widehat{u}_{T-L} \\ \widehat{y}_0 & \widehat{y}_1 & \dots & \widehat{y}_{T-L} \\ \vdots & \vdots & \dots & \vdots \\ \widehat{u}_{L-1} & \widehat{u}_{L-1} & \dots & \widehat{u}_T \\ \widehat{y}_{L-1} & \widehat{y}_{L-1} & \dots & \widehat{y}_T \end{bmatrix}. \tag{1}$$

We call $\widehat{u}$ *persistently exciting of order L* if

$$\operatorname{rank} \mathfrak{H}_L(\widehat{u}) = \begin{bmatrix} \widehat{u}_0 & \widehat{u}_1 & \dots & \widehat{u}_{T-L} \\ \vdots & \vdots & \dots & \vdots \\ \widehat{u}_{L-1} & \widehat{u}_{L-1} & \dots & \widehat{u}_T \end{bmatrix} = Lm .$$

In the statement of the fundamental lemma we use the notion of controllability (see the definition on p. 327 in [1]) and that of state cardinality (see the definition on p. 326 in [1]).

Theorem 1. *Assume that the system $\mathfrak{B}$ producing the data $\left\{\begin{bmatrix}\widehat{u}_k\\ \widehat{y}_k\end{bmatrix}\right\}_{k=0,\dots,T}$ is controllable, and denote by n its state cardinality. Define*

$$\mathcal{V} := \left\{ v \in \mathbb{R}^{L(m+p)} \mid \exists \begin{bmatrix} u \\ y \end{bmatrix} \in \mathfrak{B} \text{ and } k \in \mathbb{N} \right. \\ \left. \text{s.t. } v^\top = \begin{bmatrix} u_0^\top & y_0^\top & \dots & u_{L-1}^\top & y_{L-1}^\top \end{bmatrix} \right\}.$$

If $\widehat{u}$ is persistently exciting of order $L + n$, then $\operatorname{im} \mathfrak{H}_L(\widehat{u},\widehat{y}) = \mathcal{V}$.

Proof. See Theorem 1 p. 327 in [1]. □

Theorem 1 provides a parametrization of *all* restrictions of trajectories of $\mathfrak{B}$ to the interval $[0, L-1] \cap \mathbb{N}$ in terms of the shifts of the restrictions of *one* (sufficiently persistently exciting) trajectory $\operatorname{col}(\widehat{u},\widehat{y})$ to $[0, T] \cap \mathbb{N}$. This parametrization is used to generate i-o trajectories of the system directly from data, without knowledge of a system model. The introduction of [2] outlines some advantages of a data-driven solution to simulation as compared to the two-stage process of first identifying a model for the system, and then using the model for generating trajectories. We also remark that data-driven simulation has important applications in control (see [2,3], and some of the more recent data-driven control literature, that relies on this crucial result).

In this paper we provide a parametrization analogous to that of Theorem 1 for linear, *continuous-time* invariant systems. To this purpose we identify system trajectories with the sequences of their coefficients in the Chebyshev orthogonal polynomial series representation (in the following called for brevity's sake the

* Corresponding author.
E-mail addresses: pr3@ecs.soton.ac.uk (P. Rapisarda), m.k.camlibel@rug.nl (M.K. Çamlibel), h.j.van.waarde@rug.nl (H.J. van Waarde).

https://doi.org/10.1016/j.sysconle.2023.105603

Chebyshev representation). It is well known (see section 2.4.2 of [4]) that the Chebyshev representation of the derivative of a function can be computed *directly from the Chebyshev representation of the function itself*, by matrix multiplication with a known *differentiation operator*. Using this property, we associate to a given i-o trajectory a matrix whose rows are the Chebyshev representation of the trajectory and those of a finite number of its derivatives. Such matrix plays an analogous role to that of the Hankel matrix (1), in the sense that if the input trajectory is sufficiently *persistently exciting* (according to the definition in [5]), then the linear combinations of its columns generate *all* possible Chebyshev representations of system trajectories and a finite number of their derivatives. We use this continuous-time version of the fundamental lemma to compute the output trajectory corresponding to a given input trajectory and initial conditions. Only in a couple of recent publications concerned with stochastic systems (see [6,7]) have orthogonal basis concepts (polynomial chaos expansions in that case) been used to solve data-driven control problems. We believe that approximation theory concepts and tools have great potential also for the study of *deterministic* continuous-time systems.

The structure of the paper is as follows. In Section 2 we review some basic material about Chebyshev polynomials, and in Section 3 we recall the results of [5] on continuous-time persistency of excitation and its characterization. The continuous-time version of the fundamental lemma is stated in Section 4; two important consequences thereof are illustrated in Section 5. Section 6 contains some applications of our results to data-driven simulation. In Section 7 we discuss some of the practical issues arising in implementing our approach, and we compare our results with the literature in data-driven control.

We state our results using standard systems and control concepts, but some of the proofs require familiarity with the behavioral approach; we refer the interested reader to [8,9] as suitable introductions.

Notation

We denote by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ respectively the set of natural, real and complex numbers. We denote by $\mathbb{R}[s]$ the ring of polynomials with real coefficients, and by $\mathbb{R}^{g\times q}[s]$ the set of $g \times q$ matrices with entries in $\mathbb{R}[s]$. If $p \in \mathbb{R}[s]$, then $\deg(p)$ denotes the degree of p. We associate polynomials and differential operators with constant coefficients as follows: if $p_0 + \cdots + p_L s^L \in \mathbb{R}[s]$, then we define $p\left(\frac{d}{dt}\right)$ by $p\left(\frac{d}{dt}\right) := p_0 + \cdots + p_L \frac{d^L}{dt^L}$. The notion of degree extends in a natural way to polynomial differential operators as the largest order of differentiation.

$\mathbb{R}^n$, respectively $\mathbb{C}^n$, denote the space of n-dimensional vectors with real, respectively complex, entries. $\mathbb{R}^{n\times m}$ denotes the set of $n \times m$ matrices with real entries; $\mathbb{R}^{n\times\infty}$ the set of real matrices with n rows and an infinite number of columns; and $\mathbb{R}^{\infty\times\infty}$ the set of real matrices with an infinite number of rows and columns. The transpose of a matrix M is denoted by $M^\top$, and its pseudoinverse by $M^\dagger$. If A and B are two matrices with the same number of columns, we define $\mathrm{col}(A,B) := \begin{bmatrix} A^\top & B^\top \end{bmatrix}^\top$. Given a matrix M, $\mathrm{im}\ M$ denotes its image; left $\ker M := \{v \mid v^\top M = 0\}$, the space of its left annihilators; and row space M the subspace spanned by its rows. The ith row of a matrix M is denoted by $M(i,:)$; the submatrix consisting of the rows of M from the ith to the jth is denoted $M(i:j,:)$.

$\mathcal{L}_2(\mathbb{I},\mathbb{R})$ denotes the space of square-integrable real-valued functions defined on a finite interval $\mathbb{I} := [t_0, t_1] \subset \mathbb{R}$, equipped with the standard inner product $\langle f, g\rangle = \int_{t_0}^{t_1} f(t)g(t)dt$.

2. Chebyshev polynomial orthogonal bases

Chebyshev polynomials are widely used due to their versatility (differently from Fourier representations, which can only be used to represent periodic functions); to the efficiency with which series representations can be computed (via the FFT; this is not possible e.g., for Legendre polynomials); and to their "near best" approximation property. For details, see respectively chapters 1 and 2 of [10] and section 2 in chapter 1 of [11]; and chapter 16 of [10].

2.1. Fundamental definitions

Define $\mathbb{I} := [-1, 1]$. The Chebyshev polynomials on $\mathbb{I}$[1] are $C_0(t) := 1$, $C_1(t) := t$, and $C_{n+1}(t) = 2tC_n(t) - C_{n-1}(t)$, $n \geq 2$. Denote by w the *Chebyshev weight function* $w : (-1,1) \to \mathbb{R}$ defined by $w(t) := \frac{1}{\sqrt{1-t^2}}$, $t \in (-1,1)$; the polynomials C_k are orthogonal to each other with respect to the inner product on $\mathcal{L}_2(\mathbb{I},\mathbb{R})$ defined by $\langle f, g\rangle_w := \int_{\mathbb{I}} f(t)g(t)w(t)dt$. They form a complete basis for $\mathcal{L}_2(\mathbb{I},\mathbb{R})$, equivalently their span is dense in $\mathcal{L}_2(\mathbb{I},\mathbb{R})$. Consequently, for every $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$ the sequence $\left\{\sum_{k=0}^N \widetilde{f}_k C_k\right\}_{N\in\mathbb{N}}$, where $\widetilde{f}_k := \langle f, C_k\rangle_w \in \mathbb{R}$, $k \in \mathbb{N}$ is uniquely determined by f, converges in the mean to f. If f is Lipschitz continuous, then the sequence $\left\{\sum_{k=0}^N \widetilde{f}_k C_k\right\}_{N\in\mathbb{N}}$ converges absolutely and uniformly (see Theorem 3.1 p. 17 in [10]). Moreover (this follows from the Bessel equality, see Theorem 23 in section 6 of [12]) the sequence of coefficients $\{\widetilde{f}_k\}_{k\in\mathbb{N}}$ is square-summable. Denoting

$$\widetilde{f} := \begin{bmatrix} \widetilde{f}_0 & \widetilde{f}_1 & \ldots \end{bmatrix} \text{ and } \mathfrak{C} := \begin{bmatrix} C_0 & C_1 & \ldots \end{bmatrix}^\top, \tag{2}$$

we write

$$f = \sum_{k=0}^{\infty} \widetilde{f}_k C_k = \widetilde{f}\mathfrak{C}. \tag{3}$$

We call the right-hand side of (3) the *Chebyshev basis representation* of f. We define the bijective projection $\Pi : \mathcal{L}_2(\mathbb{I},\mathbb{R}) \to \ell_2(\mathbb{N},\mathbb{R})$ by:

$$\Pi(f) := \widetilde{f}. \tag{4}$$

If $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R}^n)$ is a vector function, we denote by f_i, $i = 1,\ldots,n$ the ith component of f. If $f_i = \sum_{k=0}^\infty \widetilde{f}_{i,k} C_k$ is the Chebyshev polynomial basis representation of f_i, $i = 1,\ldots,n$, then we write

$$f = \underbrace{\begin{bmatrix} \widetilde{f}_{1,0} & \widetilde{f}_{1,1} & \ldots \\ \vdots & \vdots & \ldots \\ \widetilde{f}_{n,0} & \widetilde{f}_{n,1} & \ldots \end{bmatrix}}_{=:\widetilde{f}} \mathfrak{C}. \tag{5}$$

In the multivariable case, the projection analogous to (4) is the map $\Pi : \mathcal{L}_2(\mathbb{I},\mathbb{R}^n) \to \ell_2(\mathbb{N},\mathbb{R}^n)$ defined by $\Pi(f) := \widetilde{f}$, with $\widetilde{f}$ defined by (5); that is, the kth coefficient of $\Pi(f)$ is the n-dimensional vector $\begin{bmatrix} f_{1,k} & \ldots & f_{n,k} \end{bmatrix}^\top$.

See Comments 2 and 6 in Section 7 for details on the computation of the Chebyshev coefficients.

[1] *Shifted* Chebyshev polynomials can be defined for general intervals (t_0, t_1) with $t_0, t_1 \in \mathbb{R}$ by the transformation $t \to \frac{2}{t_1 - t_0} t - \frac{t_1 + t_0}{t_1 - t_0}$.

2.2. N-Projection

Let $f \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$ be represented by (3); we define $\pi_N(f)$ by

$$\pi_N(f) := \sum_{k=0}^{N} \widetilde{f}_k C_k \ , \tag{6}$$

and we call it the *N-truncation* or *N-projection* of the Chebyshev representation of f. The projection of a vector-valued function is defined in the natural way.

Since the sequence of N-projections $\{\pi_N(f)\}_{N\in\mathbb{N}}$ converges, the *approximation error*

$$f - \pi_N(f) = \sum_{k=N+1}^{\infty} \widetilde{f}_k C_k \ , \tag{7}$$

decays with N. It can be shown that the "more differentiable" f is, the faster (7) goes to zero with N; consequently, "well-behaved" functions can be represented up to machine precision by truncated series. See Comment 3 in Section 7 for more details about the relation of smoothness and accuracy of the projection error.

2.3. Differentiation

Since $C_k \in \mathbb{R}[t]$, also $\frac{d}{dt}C_k \in \mathbb{R}[t]$, and there exist $d_{k,j} \in \mathbb{R}$ such that

$$\frac{d}{dt}C_k = \sum_{j=0}^{\infty} d_{k,j} C_j \ , \ k \in \mathbb{N}. \tag{8}$$

Using formula (2.4.22) p. 87 of [4], it can be proved that $d_{0,k} = 0$ for all $k \in \mathbb{N}$; if ℓ is even, then $d_{\ell,k} = 2\ell$ if $k < \ell$ is even, 0 otherwise; and if ℓ is odd, then $d_{\ell,0} = \ell$, $d_{\ell,k} = 2\ell$ if $k \leq \ell - 1$ is even, and $d_{\ell,k} = 0$ otherwise. From these expressions for $d_{k,j}$ we define the infinite matrix $\mathcal{D} := \left[d_{k,j}\right]_{k,j\in\mathbb{N}}$:

$$\mathcal{D} := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 4 & 0 & 0 & 0 & \dots \\ 3 & 0 & 6 & 0 & 0 & \dots \\ 0 & 8 & 0 & 8 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{9}$$

The nonzero entries of $\mathcal{D}$ increase linearly with their indices. Define

$$\frac{d}{dt}\mathfrak{C} := \begin{bmatrix} \frac{d}{dt}C_0 & \frac{d}{dt}C_1 & \dots \end{bmatrix}^\top . \tag{10}$$

From (8) and (10) it follows that $\frac{d}{dt}\mathfrak{C} = \mathcal{D}\mathfrak{C}$. Let $f = \sum_{k=0}^{\infty} \widetilde{f}_k C_k$; assume that $\frac{d}{dt}f \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$. Then $\frac{d}{dt}f = \sum_{k=0}^{\infty} \widetilde{f}_k \ \frac{d}{dt}C_k$ and consequently

$$\frac{d}{dt}f = \widetilde{f}\frac{d}{dt}\mathfrak{C} = \widetilde{f}\mathcal{D}\mathfrak{C} = \underbrace{\widetilde{f}\mathcal{D}}_{=:\widetilde{f}^{(1)}}\mathfrak{C}. \tag{11}$$

Eq. (11) justifies the terminology *differentiation operator* for $\mathcal{D}$. Indeed, the map $\frac{d}{dt}$ on $\mathcal{L}_2(\mathbb{I}, \mathbb{R})$ induces the map $\mathcal{D} : \ell_2(\mathbb{N}, \mathbb{R}) \to \ell_2(\mathbb{N}, \mathbb{R})$ defined by:

$$\mathcal{D}(\widetilde{f}) := \widetilde{f}\mathcal{D}. \tag{12}$$

It is straightforward to check that if $\frac{d^k}{dt^k}f \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$, then the Chebyshev representation of $\frac{d^k}{dt^k}f$ is associated with $\widetilde{f}^{(k)} := \widetilde{f}\mathcal{D}^k$, $k \geq 0$. With the position (5), the derivative of a vector-valued function satisfies the same equation as (11).

It follows from (11) that if $\widetilde{f}_k$ decays much faster than the entries of $\mathcal{D}$ (linearly) increase, then for practical purposes the computation of the coefficients $\widetilde{f}_k^{(1)}$ can be performed via multiplication of *finite* vectors and matrices: a machine-precision approximation of the (Chebyshev representation of the) derivative of a function can be *directly* computed from the (Chebyshev representation of the) function itself. See Comment 5 in Section 7 for a discussion of error bounds for the derivative approximation error.

3. Persistency of excitation

We summarize the main result of [5]. Consider the input-state-output representation

$$\begin{aligned} \frac{d}{dt}x &= Ax + Bu \\ y &= Cx + Du \ , \end{aligned} \tag{13}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $D \in \mathbb{R}^{p\times m}$. We associate to (13) its *external behavior*, defined by

$$\begin{aligned} \mathfrak{B} = \Big\{&\mathrm{col}(u, y) \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^{m+p}) \mid \exists\, x \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R}^n) \\ &\text{s.t. } \mathrm{col}(u, x, y) \text{ satisfies (13)}\Big\} . \end{aligned} \tag{14}$$

Remark 1. Other spaces than $\mathcal{C}^\infty$ could be adopted for the solution space, for example locally integrable functions, to accommodate functions relevant for engineering purposes such as ramps and steps (see [9]). We choose $\mathcal{C}^\infty$ to reduce the level of detail necessary to discuss technicalities related to convergence issues and error bounds. Making the necessary alterations, however, most of the results we present hold e.g., for Lipschitz continuous, piecewise continuous and $\mathcal{L}_2$-functions with a finite number of derivatives in the distributional sense. Moreover, using *piecewise* Chebyshev representations, also discontinuous functions such as steps can be accommodated (see section 8.1 of [13]). □

Define

$$\mathcal{O}_k := \begin{cases} C & \text{if } k = 0 \\ \begin{bmatrix} \mathcal{O}_{k-1} \\ CA^k \end{bmatrix} & \text{if } k \geq 1 \end{cases} ;$$

the *system lag* is defined by

$$\ell := \min\{k \in \mathbb{N} \mid \mathrm{rank}\ \mathcal{O}_k = \mathrm{rank}\ \mathcal{O}_{k-1}\}. \tag{15}$$

Evidently $\ell \leq n$; if (C, A) is observable, then ℓ is its observability index.

The *module of annihilators* of $\mathfrak{B}$, denoted $\mathcal{N}(\mathfrak{B})$, is the set of all polynomial differential operators annihilating all trajectories of $\mathfrak{B}$; it is formally defined by

$$\begin{aligned} \mathcal{N}(\mathfrak{B}) := \Bigg\{ &\begin{bmatrix} \eta & \xi \end{bmatrix} \in \mathbb{R}^{1\times(m+p)}[s] \mid \\ &\eta\left(\frac{d}{dt}\right)u + \xi\left(\frac{d}{dt}\right)y = 0\ \forall\ \mathrm{col}(u, y) \in \mathfrak{B} \Bigg\} . \end{aligned} \tag{16}$$

We denote by $\mathcal{N}(\mathfrak{B})^L$ the set of annihilators of $\mathfrak{B}$ with order less than or equal to L:

$$\mathcal{N}(\mathfrak{B})^L := \{\begin{bmatrix} \eta & \xi \end{bmatrix} \in \mathcal{N}(\mathfrak{B}) \mid \deg \begin{bmatrix} \eta & \xi \end{bmatrix} \leq L\} ,$$

and by $\langle \mathcal{N}(\mathfrak{B})^L \rangle$ the module of $\mathbb{R}^{1\times(m+p)}[s]$ generated by $\mathcal{N}(\mathfrak{B})^L$.

The following is Definition 1 in [5].

Definition 1 (*Persistency of Excitation*). Let $\mathbb{I} = [-1, 1]$. $f : \mathbb{I} \to \mathbb{R}^m$ is *persistently exciting of order k* if

(a) f is $(k-1)$-times continuously differentiable in $\mathbb{I}$;

(b) For every $v := \begin{bmatrix} v_0 & \dots & v_{k-1} \end{bmatrix} \in \mathbb{R}^{1\times km}$ it holds that

$$v \begin{bmatrix} f(t) \\ f^{(1)}(t) \\ \vdots \\ f^{(k-1)}(t) \end{bmatrix} = 0 \quad \text{for every } t \in \mathbb{I} \implies v_0, \dots, v_{k-1} = 0. \tag{17}$$

Remark 2. If f is persistently exciting of order k, then it does not satisfy any linear, constant-coefficient differential equation of order less than or equal to $k-1$. In the scalar case, any linear combination of k linearly independent exponential functions is persistently exciting of order k on every interval. □

For other definitions of persistency of excitation for continuous-time systems, see Def. 3.2 and Th. 3.1 in [14] in the context of adaptive parameter estimation; and [15] for the case of linear and nonlinear autonomous systems. A thorough investigation of the relation between these definitions and Definition 1 is a matter for future research.

The following is the main result of [5], see Corollary 3 p. 592 therein.

Theorem 2. *Let $(\widehat{u}, \widehat{x}, \widehat{y}) : \mathbb{R} \to \mathbb{R}^{m+n+p}$ be a trajectory of (13). Assume that (A, B) is controllable and that $\widehat{u}$ is persistently exciting of order at least $\ell + n$ on $\mathbb{I}$. Define*

$$\mathcal{N}(\widehat{u}, \widehat{y})^{\ell+1} := \left\{ \begin{bmatrix} \eta & \xi \end{bmatrix} \in \mathbb{R}^{1+(m+p)}[s] \mid \deg \begin{bmatrix} \eta & \xi \end{bmatrix} \leq \ell \right. \left. \text{and } \eta\left(\frac{d}{dt}\right)\widehat{u} + \xi\left(\frac{d}{dt}\right)\widehat{y} = 0 \text{ on } \mathbb{I} \right\}.$$

Then $\langle \mathcal{N}(\widehat{u}, \widehat{y})^{\ell+1} \rangle = \mathcal{N}(\mathfrak{B})$.

In Theorem 2 it is stated that if (A, B) is controllable and if $\widehat{u}$ is persistently exciting of order at least $n + \ell$, then the set of *all* differential equations satisfied by *all* i-o trajectories of $\mathfrak{B}$ coincides with the set of all differential equations *of order up to* ℓ satisfied by the *particular* trajectory $(\widehat{u}, \widehat{y})$. Consequently, if $\widehat{u}$ is sufficiently persistently exciting, then the data $(\widehat{u}, \widehat{y})$ are *sufficiently informative*: they contain all information about the system dynamics, as represented by the differential equations describing the system behavior. The result of Theorem 2, however, does not clarify how to exploit the sufficient informativity property of the data to *generate* system trajectories (as happens in the discrete-time case, see Theorem 1). We consider this fundamental problem in the next section.

4. An approximate fundamental lemma

Because of linearity and time-invariance of the representation (13), independently of the stability of the system none of the trajectories of $\mathfrak{B}$ defined by (14) exhibits finite escape to infinity. Consequently, for fixed $\mathbb{I}$ such trajectories all belong to $\mathcal{L}_2(\mathbb{I}, \mathbb{R}^{m+p})$, and have a Chebyshev representation. Using (4), we define

$$\Pi(\mathfrak{B}) := \left\{ \text{col}(\widetilde{u}, \widetilde{y}) \in \ell_2(\mathbb{N}, \mathbb{R}^{m+p}) \mid \exists \ \text{col}(u, y) \in \mathfrak{B} \right. \left. \text{s.t. } \text{col}(\widetilde{u}, \widetilde{y}) = \Pi(\text{col}(u, y)) \right\}. \tag{18}$$

Consider $\eta \in \mathbb{R}^{1\times(m+p)}[s]$; write $\eta(s) =: \sum_{j=0}^{\delta} \begin{bmatrix} \eta_j^u & \eta_j^y \end{bmatrix} s^j$, where $\eta_j^u \in \mathbb{R}^{1\times m}$, $\eta_j^y \in \mathbb{R}^{1\times p}$, $j = 0, \dots, \delta$. Let $L \in \mathbb{N}$, $L > \delta$; associate bijectively with $\eta(s)$ and with $\eta\left(\frac{d}{dt}\right)$ the *coefficient vector* $\widetilde{\eta} \in \mathbb{R}^{1\times L(m+p)}$ defined by

$$\widetilde{\eta} := \begin{bmatrix} \eta_0^u & \eta_0^y & \dots & \eta_\delta^u & \eta_\delta^y & 0_{1\times(m+p)} & \dots & 0_{1\times(m+p)} \end{bmatrix}. \tag{19}$$

Lemma 1. *Define $\mathfrak{B}$ by (14), and let $\text{col}(u, y) \in \mathfrak{B}$. Assume that (A, B) is controllable, and that u is persistently exciting of order $L > \ell + n$. Define the $L(m+p) \times \infty$ matrix $\mathcal{W}_L(\widetilde{u}, \widetilde{y})$ by*

$$\mathcal{W}_L(\widetilde{u}, \widetilde{y}) := \text{col}\left(\begin{bmatrix} \widetilde{u} \\ \widetilde{y} \end{bmatrix} \mathcal{D}^j \right)_{j=0,\dots,L-1}. \tag{20}$$

Then

$$\textit{left ker } \mathcal{W}_L(\widetilde{u}, \widetilde{y}) = \left\{ \widetilde{\eta} \in \mathbb{R}^{L(m+p)} \mid \eta\left(\frac{d}{dt}\right) \in \mathcal{N}(\mathfrak{B}) \right\}. \tag{21}$$

Proof. We use a *row-proper* kernel representation of $\mathfrak{B}$ (see p. 576 of [16]) $R(s) = R_0 + \dots + R_\ell$, with $R_i \in \mathbb{R}^{p\times(m+p)}$, $i = 0, \dots, \ell$. Denote the ith row of $R(s)$ by $\eta_i(s)$ and its degree by δ_i, $i = 1, \dots, p$. It can be proved (see Corollary 6.7 p. 1065 of [9]) that $n = \sum_{i=1}^{p} \delta_i$.

Write $\eta_i(s) =: \sum_{j=0}^{\delta_i} \begin{bmatrix} \eta_{i,j}^u & \eta_{i,j}^y \end{bmatrix} s^j$. Since $\eta_i\left(\frac{d}{dt}\right)\text{col}(u', y') = 0$ for every $\text{col}(u', y') \in \mathfrak{B}$, and since $L > \delta_i$, the corresponding coefficient vector $\widetilde{\eta}_i$ defined by (19) belongs to left ker $\mathcal{W}_L$, $i = 1, \dots, p$. Since $\frac{d^k}{dt^k}\eta_i\left(\frac{d}{dt}\right)$, $k \in \mathbb{N}$ also annihilates $\mathfrak{B}$, it follows that the *shifts* of $\widetilde{\eta}_i$, defined by

$$\sigma\widetilde{\eta}_i := \begin{bmatrix} 0 & 0 & \eta_{i,0}^u & \eta_{i,0}^y & \dots & \eta_{i,\delta_i}^u & \eta_{i,\delta_i}^y & 0 & 0 & \dots \end{bmatrix}$$
$$\sigma^2\widetilde{\eta}_i := \begin{bmatrix} 0 & 0 & 0 & 0 & \eta_{i,0}^u & \eta_{i,0}^y & \dots & \eta_{i,\delta_i}^u & \eta_{i,\delta_i}^y & \dots \end{bmatrix}$$
$$\vdots$$
$$\sigma^{L-\delta_i-1}\widetilde{\eta}_i := \begin{bmatrix} 0 & 0 & \dots & \eta_{i,0}^u & \eta_{i,0}^y & \dots & \eta_{i,\delta_i}^u & \eta_{i,\delta_i}^y \end{bmatrix}$$

also belong to the left kernel of $\mathcal{W}_L$. The association of $\frac{d^k}{dt^k}\eta_i\left(\frac{d}{dt}\right)$ and $\sigma^k\widetilde{\eta}_i$ defines an isomorphism between polynomial differential operators and their coefficient vectors.

Since $R(s)$ is row-proper, its highest row-coefficient matrix R_{hc}, defined by $R_{\text{hc}} := \text{col}(\begin{bmatrix} \eta_{i,\delta_i}^u & \eta_{i,\delta_i}^y \end{bmatrix})_{i=1,\dots,p}$, has full row rank. Consequently, the vectors $\sigma^j\widetilde{\eta}_i$, $j = 0, \dots, L-\delta_i-1$, $i = 1, \dots, p$ are linearly independent. Note that there are $\sum_{i=1}^{p}(L-\delta_i) = Lp - \sum_{i=1}^{p}\delta_i = Lp - n$ such vectors. Define the $(Lp-n) \times L(m+p)$ matrix

$$E := \text{col}\left(\sigma^j\widetilde{\eta}_i\right)_{i=1,\dots,p,j=0,\dots,L-\delta_i-1}; \tag{22}$$

since $\mathcal{N}(\mathfrak{B}) = \langle \eta_1(s), \dots, \eta_p(s) \rangle$, the equality

$$\text{row space}(E) = \left\{ \widetilde{\eta} \in \mathbb{R}^{L(m+p)} \mid \eta\left(\frac{d}{dt}\right) \in \mathcal{N}(\mathfrak{B}) \right\},$$

holds true. Since $\text{col}(u, y) \in \mathfrak{B}$, for $i = 1, \dots, p$ it holds that $\eta_i\left(\frac{d}{dt}\right)\text{col}(u, y) = \eta_i\left(\frac{d}{dt}\right)\sum_{k=0}^{\infty}\text{col}(\widetilde{u}_k, \widetilde{y}_k)C_k = 0$ and from the linear independence of the Chebyshev polynomials C_k it follows that $E\mathcal{W}_L(\widetilde{u}, \widetilde{y}) = 0$. Consequently, row space(E) is contained in left ker $(\mathcal{W}_L(\widetilde{u}, \widetilde{y}))$.

We prove the converse inclusion. Recall that u is persistently exciting of order greater than $\ell + n$. Denote by $\mathcal{N}(u, y)^L$ the set of annihilators of $\text{col}(u, y)$ with degree less than or equal to $L-1$, and by $\langle \mathcal{N}(u, y)^L \rangle$ the module of $\mathbb{R}^{1\times(m+p)}[s]$ generated by these vectors. Use Theorem 2 to conclude that since $L > \ell + n$, the equality $\mathcal{N}(\mathfrak{B}) = \langle \mathcal{N}(u, y)^L \rangle$ holds. Consequently, $\mathcal{N}(\mathfrak{B}) = \langle \eta_1(s), \dots, \eta_p(s) \rangle = \langle \mathcal{N}(u, y)^L \rangle$. Each $\eta(s) \in \mathcal{N}(u, y)^L$ is associated with a vector $\widetilde{\eta} \in \mathbb{R}^{1\times L(m+p)}$ that belongs to the left kernel of $\mathcal{W}_L(\widetilde{u}, \widetilde{y})$. It follows that row space$(E) \supseteq$ left ker $(\mathcal{W}_L(\widetilde{u}, \widetilde{y}))$. □

Corollary 1. *Define $\mathfrak{B}$ by (14), and let $\text{col}(u, y) \in \mathfrak{B}$. Assume that (A, B) is controllable, and that u is persistently exciting of order $L > \ell + n$. Define $\mathcal{W}_L(\widetilde{u}, \widetilde{y})$ by (20). Then rank $\mathcal{W}_L(\widetilde{u}, \widetilde{y}) = Lm + n$.*

Proof. Define the annihilators $\eta_i(s)$ as in the proof of Lemma 1. From Lemma 1 it follows that the subspace of $\mathbb{R}^{L(m+p)}$ generated by the rows of E defined by (22) equals left ker $\mathcal{W}_L(\widetilde{u}, \widetilde{y})$.

Since R is row-proper, E has full row rank $Lp-n$. Consequently rank $\mathcal{W}_L(\widetilde{u},\widetilde{y}) = L(m+p)-(Lp-n) = Lm+n$. $\square$

We state the continuous-time equivalent of Theorem 1.

Theorem 3. *Define $\mathfrak{B}$ by (14), and let $\operatorname{col}(u,y) \in \mathfrak{B}$. Assume that (A,B) is controllable, and that u is persistently exciting of order $L > \ell + n$. Define $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ by (20), and*

$$\mathcal{V} := \left\{ v \in \mathbb{R}^{L(m+p)} \mid \exists \ \operatorname{col}(\overline{u},\overline{y}) \in \mathfrak{B}\ ,\ k \in \mathbb{N} \right.$$
$$\left. \text{s.t. } \operatorname{col}\left(\begin{bmatrix} \widetilde{\overline{u}}_k^{(j)} \\ \widetilde{\overline{y}}_k^{(j)} \end{bmatrix}\right)_{j=0,\dots,L-1} = v \right\}. \tag{23}$$

Then im $\mathcal{W}_L(\widetilde{u},\widetilde{y}) = \mathcal{V}$.

Proof. The subspace im $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ is included in $\mathcal{V}$, since it consists of finite linear combinations of columns $\operatorname{col}\left(\begin{bmatrix} \widetilde{u}_k^{(j)} \\ \widetilde{y}_k^{(j)} \end{bmatrix}\right)_{j=0,\dots,L-1}$, and since $\operatorname{col}(u,y) \in \mathfrak{B}$.

We prove the converse inclusion. Since (A,B) is controllable, $\mathfrak{B}$ admits an image representation (see section 6.6 of [8]) induced by a polynomial matrix $M \in \mathbb{R}^{(m+p)\times m}[s]$. That is, $\mathfrak{B} = \operatorname{im} M\left(\frac{d}{dt}\right)$, where we write

$$M(s) =: \begin{bmatrix} D(s) \\ N(s) \end{bmatrix} =: \begin{bmatrix} D_0 + D_1 s + \cdots + D_r s^r \\ N_0 + N_1 s + \cdots + N_r s^r \end{bmatrix},$$

where $D_i \in \mathbb{R}^{m\times m}$, $N_i \in \mathbb{R}^{p\times m}$, $i = 1,\dots,r$. A trajectory $\operatorname{col}(u',y') \in \mathfrak{B}$ if and only if there exists $g : \mathbb{R} \to \mathbb{R}^m$ such that

$$\begin{bmatrix} u' \\ y' \end{bmatrix} = M\left(\frac{d}{dt}\right) g = \begin{bmatrix} D_0 & \dots & D_r \\ N_0 & \dots & N_r \end{bmatrix} \begin{bmatrix} g \\ \frac{d}{dt} g \\ \vdots \\ \frac{d^r}{dt^r} g \end{bmatrix}.$$

Define $\widetilde{D} := \begin{bmatrix} D_0 & \dots & D_r \end{bmatrix}$, $\widetilde{N} := \begin{bmatrix} N_0 & \dots & N_r \end{bmatrix}$; and define their j-shifts, $j = 0,\dots,L-1$, by

$$\sigma^j \widetilde{D} := \begin{bmatrix} 0 & \dots & 0 & D_0 & \dots & D_r & 0 & \dots & 0 \end{bmatrix}$$
$$\sigma^j \widetilde{N} := \begin{bmatrix} 0 & \dots & 0 & N_0 & \dots & N_r & 0 & \dots & 0 \end{bmatrix},$$

where the first jm columns of $\sigma^j\widetilde{D}$ and $\sigma^j\widetilde{N}$ are zero, $j = 0,\dots,L-1$.

We define $\mathcal{M} := \operatorname{col}\left(\begin{bmatrix} \sigma^j\widetilde{D} \\ \sigma^j\widetilde{N} \end{bmatrix}\right)_{j=0,\dots,L-1}$, and we write

$$\operatorname{col}\left(\begin{bmatrix} u'^{(j)} \\ y'^{(j)} \end{bmatrix}\right)_{j=0,\dots,L} = \mathcal{M} \operatorname{col}\left(g^{(j)}\right)_{j=0,\dots,L+r-1}. \tag{24}$$

Applying Π to both sides of (24) we obtain

$$\operatorname{col}\left(\begin{bmatrix} \widetilde{u}'^{(j)} \\ \widetilde{y}'^{(j)} \end{bmatrix}\right)_{j=0,\dots,L} = \mathcal{M} \operatorname{col}\left(\widetilde{g}^{(j)}\right)_{j=0,\dots,L+r-1}. \tag{25}$$

Since $\widetilde{g}$ can take the value of any sequence in $\ell_2(\mathbb{N},\mathbb{R}^m)$, the subspace $\mathcal{V}$ defined in (23) satisfies $\mathcal{V} = \operatorname{im}\mathcal{M}$. We now prove that $\operatorname{im}\mathcal{M} = \operatorname{im}\mathcal{W}_L(\widetilde{u},\widetilde{y})$.

The set of annihilators $\mathcal{N}(\mathfrak{B})$ coincides with the syzygy of the module generated by the columns of $M(s)$ (see Theorem 2.1 in [9]). The argument used in proving Lemma 1 shows that $\mathcal{N}(\mathfrak{B})$ coincides with the module generated by the polynomial vectors associated with the elements of left ker $\mathcal{W}_L(\widetilde{u},\widetilde{y})$. It follows that left ker $\mathcal{W}_L(\widetilde{u},\widetilde{y}) = (\operatorname{im}\mathcal{M})^\perp = (\mathcal{V})^\perp$, from which we conclude that $\mathcal{V} = \operatorname{im}\mathcal{W}_L(\widetilde{u},\widetilde{y})$. This concludes the proof. $\square$

Remark 3. We illustrate the conceptual relation between Theorems 1 and 3. Time-shift (in discrete-time) corresponds to differentiation in the Chebyshev representation domain (see (12)). The restrictions

$$\begin{bmatrix} u_0^\top & y_0^\top & \dots & u_{L-1}^\top & y_{L-1}^\top \end{bmatrix}^\top \tag{26}$$

of discrete-time trajectories to $[0, L-1]$ correspond to the vector

$$\begin{bmatrix} \widetilde{\overline{u}}_k^\top & \widetilde{\overline{y}}_k^\top & \dots & \widetilde{\overline{u}}_k^{(L-1)\top} & \widetilde{\overline{y}}_k^{(L-1)\top} \end{bmatrix}, \tag{27}$$

constructed from the kth Chebyshev coefficients of the jth derivative of $\operatorname{col}(\overline{u},\overline{y})$, $j = 0,\dots,L-1$. Note that (27) is the Chebyshev representation of the L-jet $\begin{bmatrix} \overline{u}_k^\top & \overline{y}_k^\top & \dots & \overline{u}_k^{(L-1)\top} & \overline{y}_k^{(L-1)\top} \end{bmatrix}$, of $\operatorname{col}(\overline{u},\overline{y})$. The matrix $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ defined in (20) plays a similar role to that of the Hankel matrix (1). The image of $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ equals the set of all admissible vectors of Chebyshev coefficients (27) corresponding to some trajectory $\operatorname{col}(\overline{u},\overline{y}) \in \mathfrak{B}$, just as the image of $\mathfrak{H}_L(\widehat{u},\widehat{y})$ equals the set of all restrictions (26) of trajectories of $\mathfrak{B}$ to $[0, L-1] \cap \mathbb{N}$. $\square$

Corollary 2. *Define $\mathfrak{B}$ by (14), and let $\operatorname{col}(u,y) \in \mathfrak{B}$. Assume that (A,B) is controllable, and that u is persistently exciting of order $L > \ell + n$. Define $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ by (20). Define $d := \operatorname{rank}(\mathcal{W}_L(\widetilde{u},\widetilde{y})) = Lm+n$ and factorize $\mathcal{W}_L(\widetilde{u},\widetilde{y})$ as $\mathcal{W}_L(\widetilde{u},\widetilde{y}) = MR$, with $M \in \mathbb{R}^{L(m+p)\times d}$, $R \in \mathbb{R}^{d\times\infty}$. Then $(\operatorname{im} M)^\perp$ is isomorphic to a set of generators of $\mathcal{N}(\mathfrak{B})$.*

Proof. Follows from Lemma 1 and Theorem 3. $\square$

Example 1. Consider data generated by the behavior with $m = p = 1$ described by the differential equation $\left(\frac{d}{dt} - 1\right) y = u$. The lag $\ell = 1$ and the state cardinality $n = 1$. The trajectory

$$\begin{bmatrix} u(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} -5e^{-4t} - 4e^{-3t} - 3e^{-2t} - 2e^{-t} \\ e^{-4t} + e^{-3t} + e^{-2t} + e^{-t} \end{bmatrix},$$

belongs to $\mathfrak{B}$, and u is persistently exciting of order $L = 3 > n + \ell = 2$. A machine-precision Chebyshev representation of $\widetilde{u}$ and $\widetilde{y}$ can be computed with 23 Chebyshev coefficients using `Chebfun` (see [17] and Comments 2 and 4 in Section 7). Let $L = 3$; the corresponding matrix $\mathcal{W}_L(\widetilde{u},\widetilde{y}) \in \mathbb{R}^{6\times 23}$ has singular values $2.7428 \cdot 10^3$, 9.6540, $3.3994 \cdot 10^{-1}$, $3.0483 \cdot 10^{-3}$, $3.5851 \cdot 10^{-11}$, $2.3100 \cdot 10^{-13}$. The gap between the last two singular values and the third last one is of the order of 10^8; this numerically confirms the results of Corollary 1. Using the same notation as in the proof of Lemma 1, define $\eta := \begin{bmatrix} 1 & 0 & 0 & 1 & -1 & 0 \end{bmatrix}$ from the values of the coefficients of the polynomial differential operator defining the system. It can be verified that $\|\eta\mathcal{W}_3(\widetilde{u},\widetilde{y})\| = 4.3783 \cdot 10^{-13}$ and $\|\sigma\eta\mathcal{W}_3(\widetilde{u},\widetilde{y})\| = 7.1132 \cdot 10^{-11}$. Since $\mathfrak{B} = \operatorname{im}\begin{bmatrix} \frac{d}{dt} - 1 \\ 1 \end{bmatrix}$, using Theorem 3 we conclude that $\mathcal{V}$ defined in (23) equals

$$\operatorname{im}\begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^\top = \operatorname{im}\begin{bmatrix} V_u^\top & V_y^\top \end{bmatrix},$$

with dimension 4, as stated in Corollary 1. In this particular case, the shifts of the coefficient matrix of the image representation can be directly used to compute a basis for $\mathcal{V}$.

If $N = 9$, the corresponding matrix $\mathcal{W}_3(\widetilde{u},\widetilde{y}) \in \mathbb{R}^{6\times 10}$ has singular values $2.7386 \cdot 10^3$, 9.1684, $7.4189 \cdot 10^{-1}$, $3.0300 \cdot 10^{-1}$, $1.7760 \cdot 10^{-2}$, and $2.3464 \cdot 10^{-3}$. In such case it is not evident what rank $\mathcal{W}_3(\widetilde{u},\widetilde{y})$ is, since the singular values are relatively close together. It can be verified that $\|\eta\mathcal{W}_3(\widetilde{u},\widetilde{y})\| = 1.7562 \cdot 10^{-1}$, and consequently computing a basis for im $\mathcal{W}_3(\widetilde{u},\widetilde{y})$ in this case would not produce an accurate system of generators for the subspace $\mathcal{V}$ defined in Theorem 3. A significant drop in the magnitude of the singular values (of the order of 10^4) appears only for values of N larger than or equal to 18; only for such larger numbers of coefficients makes sense to consider data matrices as representative of the system dynamics. $\square$

5. Data-driven characterization of system trajectories

In discrete-time, every linear combination of the shifts of (the restriction of) a system trajectory is (the restriction of) a system trajectory. Consequently, $\mathfrak{H}_L(\widehat{u},\widehat{y})$ defined in (1) provides a *direct* parametrization of all system trajectories. Linearly combining the columns of $\mathcal{W}(\widetilde{u},\widetilde{y})$ defined in (20) (or the columns of M defined from a rank-revealing factorization of $\mathcal{W}(\widetilde{u},\widetilde{y})$, see Corollary 2) only generates the kth *coefficients* of the L-jet of a system trajectory. To compute from $\mathcal{W}(\widetilde{u},\widetilde{y})$ the Chebyshev representation of an admissible system trajectory, a *sequence* of $L(m+p)$-dimensional vectors needs to be computed by linear combination of its columns. In the next two subsections we show how to do this in two important cases: the computation of an i-o system trajectory; and the computation of the output trajectory corresponding to a given input one.

5.1. Data-driven characterization of all system trajectories

We first introduce some notation. In the following it will be easier to work with a rearranged version of (20) and the corresponding subspace (23). We define

$$\begin{bmatrix} \operatorname{col}\left(\left[\widetilde{u}\mathcal{D}^j\right]\right)_{j=0,\dots,L-1} \\ \operatorname{col}\left(\left[\widetilde{y}\mathcal{D}^j\right]\right)_{j=0,\dots,L-1}\end{bmatrix} =: \begin{bmatrix}\mathcal{W}_L(\widetilde{u})\\ \mathcal{W}_L(\widetilde{y})\end{bmatrix}. \tag{28}$$

and $\mathcal{V}$ by

$$\mathcal{V} := \left\{ v\in\mathbb{R}^{L(m+p)} \mid \exists\ \operatorname{col}(\overline{u},\overline{y})\in\mathfrak{B}\ ,\ k\in\mathbb{N} \right.$$
$$\left. \text{such that } \begin{bmatrix}\operatorname{col}\left(\widetilde{\overline{u}}_k^{(j)}\right)_{j=0,\dots,L-1}\\ \operatorname{col}\left(\widetilde{\overline{y}}_k^{(j)}\right)_{j=0,\dots,L-1}\end{bmatrix} = v \right\}. \tag{29}$$

Denote $d := \dim(\mathcal{V})$; recall from Corollary 1 that $d = Lm+n$. Let $V\in\mathbb{R}^{L(m+p)\times d}$ be a basis matrix for $\mathcal{V}$; such a basis matrix can be computed via a rank-revealing factorization of $\begin{bmatrix}\mathcal{W}_L(\widetilde{u})\\ \mathcal{W}_L(\widetilde{y})\end{bmatrix}$, see Corollary 2. Partition V conformably with (29):

$$V =: \begin{bmatrix}V_u\\ V_y\end{bmatrix}\ ,\ \text{with } V_u\in\mathbb{R}^{Lm\times d}\ ,\ V_y\in\mathbb{R}^{Lp\times d}. \tag{30}$$

Finally, recall that we denote the ith row of a matrix M by $M(i,:)$.

Theorem 4. *Define $\mathfrak{B}$ by (14), and let $\operatorname{col}(u,y)\in\mathfrak{B}$. Assume that (A,B) is controllable, and that u is persistently exciting of order $L>\ell+n$. Define $\mathcal{W}_L(\widetilde{u})$, $\mathcal{W}_L(\widetilde{y})$ by (28); V_u and V_y by (30); and $\Pi(\mathfrak{B})$ by (18). The following statements are equivalent:*

1. $\operatorname{col}\left(\widetilde{u}',\widetilde{y}'\right)\in\Pi(\mathfrak{B})$;
2. *There exists $G\in\mathbb{R}^{d\times\infty}$ such that*

$$\begin{bmatrix}\mathcal{W}(\widetilde{u}')\\ \mathcal{W}(\widetilde{y}')\end{bmatrix} = \begin{bmatrix}V_u\\ V_y\end{bmatrix} G\,, \tag{31}$$

3. *There exists $G\in\mathbb{R}^{d\times\infty}$ such that*

$$\begin{aligned}(V_uG)(1,:) &= \widetilde{u}'\\ (V_yG)(1,:) &= \widetilde{y}'\\ (V_uG)(1,:)\mathcal{D}^i-(V_uG)(i+1,:) &= 0\\ (V_yG)(1,:)\mathcal{D}^i-(V_yG)(i+1,:) &= 0\ ,\\ i = 1,\dots,L-1.\end{aligned} \tag{32}$$

Proof. To prove the implication (1) $\Longrightarrow$ (2), construct from the orthogonal basis representation of $\operatorname{col}(u',y')$ the matrices $\mathcal{W}(\widetilde{u}')$ and $\mathcal{W}(\widetilde{y}')$. From Theorem 3 it follows that $\operatorname{im}\begin{bmatrix}\mathcal{W}_L(\widetilde{u}')\\ \mathcal{W}_L(\widetilde{y}')\end{bmatrix}\subseteq \operatorname{im}\begin{bmatrix}\mathcal{W}_L(\widetilde{u})\\ \mathcal{W}_L(\widetilde{y})\end{bmatrix} = \operatorname{im} V$. Consequently, there exists $G\in\mathbb{R}^{d\times\infty}$ such that $\begin{bmatrix}\mathcal{W}_L(\widetilde{u}')\\ \mathcal{W}_L(\widetilde{y}')\end{bmatrix} = VG$.

We prove the implication (2) $\Longrightarrow$ (1). Let $\eta\in\mathbb{R}^{1\times(m+p)}[s]$ be any annihilator of $\mathfrak{B}$ of degree less than or equal to $L-1$; define its coefficient vector $\widetilde{\eta}$ by (19). Since V is a basis matrix for $\mathcal{V}$ defined by (29), it follows from (31) that $\widetilde{\eta}\begin{bmatrix}\mathcal{W}_L(\widetilde{u}')\\ \mathcal{W}_L(\widetilde{y}')\end{bmatrix} = 0$. Now $0 = \widetilde{\eta}\begin{bmatrix}\mathcal{W}_L(\widetilde{u}')\\ \mathcal{W}_L(\widetilde{y}')\end{bmatrix} = \sum_{j=0}^{L-1}\widetilde{\eta}_j^u\left(\widetilde{u}'\mathcal{D}^j\right)+\sum_{j=0}^{L-1}\widetilde{\eta}_j^y\left(\widetilde{y}'\mathcal{D}^j\right)$, is equivalent with $0=\sum_{j=0}^{L-1}\widetilde{\eta}_j^u\left(\frac{d^j}{dt^j}u'\right)+\sum_{j=0}^{L-1}\widetilde{\eta}_j^y\left(\frac{d^j}{dt^j}y'\right)$. It follows that $\operatorname{col}(u',y')$ is annihilated by every annihilator of $\mathfrak{B}$. It follows that $\operatorname{col}(u',y')\in\mathfrak{B}$, and consequently that $\operatorname{col}(\widetilde{u}',\widetilde{y}')\in\Pi(\mathfrak{B})$.

To prove (2) $\Longleftrightarrow$ (3), observe that the ith row of $\mathcal{W}_L(\widetilde{u}')$ equals $\widetilde{u}'\mathcal{D}^i$, $i=0,\dots,L-1$. The equivalence of $\mathcal{W}_L(\widetilde{u}')=V_uG$ and the first and third equation in (32) follows. The equivalence of $\mathcal{W}_L(\widetilde{y}')=V_yG$ with the second and last equation in (32) follows analogously. □

5.2. Data-driven characterization of system trajectories given the input

In data-driven simulation, the problem arises of computing output trajectories with specified input trajectories. We provide a refinement of the result of Theorem 4 and characterize all output trajectories corresponding to a given input. To this purpose we need to state three preliminary results.

Proposition 1. *Define the subspace $\mathcal{V}$ by (29). Let V be a basis matrix for it, and partition it as in (30). The matrix V_u has full row rank Lm.*

Proof. Since u is an input, $\operatorname{col}(u^{(i)})_{i=0,\dots,L-1}$ can be any trajectory in $\mathcal{L}_2(\mathbb{I},\mathbb{R}^{Lm})$, and its Chebyshev representation $\operatorname{col}(\widetilde{u}^{(i)})_{i=0,\dots,L-1}$ can take any value in the space of m-dimensional square-summable sequences $\ell_2(\mathbb{I},\mathbb{R}^{Lm})$.

Consequently, for every $v_u\in\mathbb{R}^{Lm}$ there exists $u'\in\mathcal{L}_2(\mathbb{I},\mathbb{R})$ and $k\in\mathbb{N}$ such that the Chebyshev representation $\operatorname{col}(\widetilde{u}'^{(i)})_{i=0,\dots,L-1}$ of $\operatorname{col}(u'^{(i)})_{i=0,\dots,L-1}$, the L-jet of u', satisfies $v_u = \operatorname{col}\left(\widetilde{u}_k'^{(i)}\right)_{i=0,\dots,L-1}$. Choose one output trajectory $y'\in\mathcal{L}_2(\mathbb{I},\mathbb{R}^p)$ corresponding to u', and consider the vector $v_y := \operatorname{col}(\widetilde{y}_k'^{(i)})_{i=0,\dots,L-1}$. Then $v := \operatorname{col}(v_u,v_y)\in\mathcal{V}$, and $v_u\in\operatorname{im}(V_u)$. Since v_u is arbitrary, V_u is surjective. □

Proposition 2. *Let V be a basis matrix for the subspace $\mathcal{V}$ defined by (29); partition it as in (30). Let K be a basis matrix for $\ker V_u$; then $\operatorname{rank} V_yK = n$, the state cardinality of $\mathfrak{B}$.*

Proof. From Proposition 1 and Corollary 1 it follows that $\dim\ \ker V_u = d-Lm = Lm+n-Lm = n$. The matrix V_yK has Lp rows and n columns. Assume that there exists $v\in\mathbb{R}^n$ such that $(V_yK)v=0$; then, since $\operatorname{im}\ K = \ker\ V_u$, it follows that $V_u(Kv)$=0. Consequently, $\begin{bmatrix}V_u\\ V_y\end{bmatrix}(Kv)=0$. The matrix V has full column rank; the equality $Kv=0$ follows. K is a basis matrix, and consequently $v=0$. Conclude that V_yK has full column rank. □

Proposition 3. *Let $u\in\mathcal{L}_2(\mathbb{I},\mathbb{R}^m)$, with Chebyshev representation $\widetilde{u}\in\ell_2(\mathbb{I},\mathbb{R}^m)$; define $\mathcal{W}_L(\widetilde{u})$ by (28). Let V be an image matrix for (29), partitioned as in (30). Let $K\in\mathbb{R}^{d\times n}$ be a basis matrix for $\ker\ V_u$. The following statements are equivalent:*

1. $G \in \mathbb{R}^{Lm \times \infty}$ *solves the equation* $V_u G = \mathcal{W}_L(\widetilde{u})$;
2. *There exists* $F \in \mathbb{R}^{n \times \infty}$ *such that*

$$G = V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}) + KF. \tag{33}$$

Proof. Straightforward from the fact that V_u is surjective. □

We characterize all output trajectories corresponding to a given input.

Theorem 5. *Define* $\mathfrak{B}$ *by* (14), *and let* $\operatorname{col}(u, y) \in \mathfrak{B}$. *Assume that* (A, B) *is controllable, and that* u *is persistently exciting of order* $L > \ell + n$. *Define* $\mathcal{W}_L(\widetilde{u})$, $\mathcal{W}_L(\widetilde{y})$ *by* (28); V_u *and* V_y *by* (30); $\Pi(\mathfrak{B})$ *by* (18); *and denote by* $K \in \mathbb{R}^{d \times n}$ *a basis matrix for* $\ker V_u$. *Let* $u' \in \mathcal{L}_2(\mathbb{I}, \mathbb{R}^m)$ *be given. The following statements are equivalent:*

1. $\operatorname{col}\left(\widetilde{u}', \widetilde{y}'\right) \in \Pi(\mathfrak{B})$;
2. *There exists* $F \in \mathbb{R}^{n \times \infty}$ *such that*

$$\mathcal{W}_L(\widetilde{y}') = V_y \left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}') + KF\right) .$$

Proof. To prove (1) $\Longrightarrow$ (2), construct from the representation of $\operatorname{col}(u', y')$ the matrices $\mathcal{W}_L(\widetilde{u}')$ and $\mathcal{W}_L(\widetilde{y}')$. Use Theorem 3 to conclude that $\begin{bmatrix} \mathcal{W}_L(\widetilde{u}') \\ \mathcal{W}_L(\widetilde{y}') \end{bmatrix} = VX$ for some $X \in \mathbb{R}^{d \times \infty}$. Use formula (33) in Proposition 3 to conclude that $X = V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}) + KF$ for some F. Statement (2) follows.

To prove (2) $\Longrightarrow$ (1), observe that if F is such that

$$\mathcal{W}_L(\widetilde{y}') = V_y \left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}(\widetilde{u}) + KF\right) ,$$

then also

$$\begin{bmatrix} \mathcal{W}_L(\widetilde{u}') \\ \mathcal{W}_L(\widetilde{y}') \end{bmatrix} = \begin{bmatrix} V_u \\ V_y \end{bmatrix} \left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}(\widetilde{u}) + KF\right). \tag{34}$$

Now use statement (2) of Theorem 4. □

6. Data-driven continuous-time simulation

Define $\mathfrak{B}$ by (14), and let $\operatorname{col}(u, y) \in \mathfrak{B}$. Assume that (A, B) is controllable, and that u is persistently exciting of order $L > \ell + n$. Define $\mathcal{W}_L(\widetilde{u})$, $\mathcal{W}_L(\widetilde{y})$ by (28); V_u and V_y by (30); and $\Pi(\mathfrak{B})$ by (18).

Problem: *Given* $q \in \mathbb{N}$; $d_i \in \mathbb{R}^{(m+p)}$, $i = 0, \ldots, q-1$; $t_0 \in \mathbb{I}$; *and* $u' \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$, *compute if it exists, a trajectory* $y' \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$ *such that* $\operatorname{col}(u', y') \in \mathfrak{B}$ *and*

$$\operatorname{col}\left(u'^{(i)}(t_0), y'^{(i)}(t_0)\right) = d_i \ , \ \ i = 0, \ldots, q-1. \tag{35}$$

In the following we call this the *simulation problem.*

Proposition 4. *Denote by* $V_y((i-1)p + 1 : ip, :)$ *the ith p-dimensional block-row of* V_y, $i = 1, \ldots, L+1$. *Let* K *be a basis matrix for* $\ker V_u$. *The simulation problem is solvable if and only if there exists* $F \in \mathbb{R}^{n \times \infty}$ *such that*

$$\begin{aligned} &V_y((i-1)p + ip, :)\left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}') + KF\right)\mathcal{D} \\ &= V_y(ip+1 : (i+1)p, :)\left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}') + KF\right) , \end{aligned} \tag{36}$$

$i = 1, \ldots, L-1$; *and*

$$\begin{bmatrix} \widetilde{u}' \\ V_y(1, :)\left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}') + KF\right) \end{bmatrix} \mathcal{D}^i \mathfrak{C}(t_0) = d_i \ , \tag{37}$$

$i = 0, \ldots, q-1$. *If the problem is solvable, then the solution is the trajectory* $\operatorname{col}(u', y')$ *whose orthogonal basis coefficients are*

$$\operatorname{col}\left(\widetilde{u}', V_y(1:p, :)\left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}_L(\widetilde{u}) + KF\right)\right) .$$

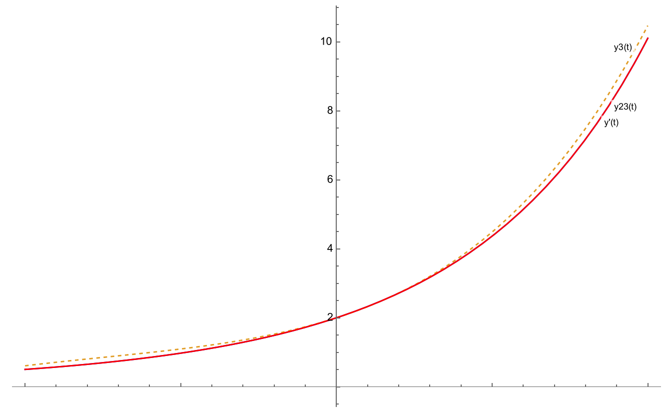

Fig. 1. Simulated and exact trajectories for Example 2.

Proof. The conditions (36) are equivalent to each block-row of $V_y F$ being the Chebyshev representation of the derivative of the vector trajectory represented by the previous block-row. The conditions (37) are equivalent with (35). □

Example 2. We consider the same data as in Example 1. We choose the input trajectory $u'(t) = e^{2t}$, and the conditions $\begin{bmatrix} u'(0) \\ y'(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Solving the differential equation for the given input and initial conditions yields the system trajectory whose value at t is $\begin{bmatrix} u'(t) \\ y'(t) \end{bmatrix} = \begin{bmatrix} e^{2t} \\ e^t + e^{2t} \end{bmatrix}$. We now compute an approximate solution directly from the data.

Since $V_u = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$ and $V_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, choosing $K = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^\top$ results in $V_y K = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$. As stated in Proposition 2, rank $V_y K$ equals the state-space dimension $n = 1$.

To approximate the input function u' to machine precision, we use 23 Chebyshev coefficients. We then solve with `Mathematica` the Eqs. (36); the solutions are parametrized by $\alpha \in \mathbb{R}$ as

$$\begin{aligned} F(\alpha) = \alpha &\left[\ 1 \quad 0.8928 \quad 0.2144 \quad 0.0350 \quad 0.0043 \quad \ldots\ \right] \\ &+\left[\ 0 \quad 4.2979 \quad 3.3340 \quad 1.2962 \quad 0.3435 \quad \ldots\ \right]. \end{aligned}$$

The Eq. (37) imposing the satisfaction of the initial condition for such family of parametrized sequences yields the solution $\alpha = 9.815 =: \overline{\alpha}$. We conclude from Proposition 4 that a solution to this data-driven simulation problem exists. The polynomial approximating such solution corresponds to the entries of

$$V_{y,1,:}\left(V_u^\top \left(V_u V_u^\top\right)^{-1} \mathcal{W}(\widetilde{u}') + KF(\overline{\alpha})\right) ,$$

see statement 2 of Theorem 5.

We computed with `Mathematica` the $\mathcal{L}_2$-norm of the error on $\mathbb{I} = [-1, 1]$ between such polynomial and the exact solution $y'(t) = e^t + e^{2t}$; it equals $2.32831 \cdot 10^{-10}$. A plot of the exact solution y' and of the simulated one y_{23} is given in Fig. 1; they are indistinguishable from each other.

To illustrate the effect of truncation on the accuracy of the simulated trajectories, we approximate u' using only 3 coefficients. Solving (36) and (37) with `Mathematica` yields

$$\begin{aligned} F'(\alpha) = \alpha &\left[\ 0 \quad 0.8928 \quad 0.2144 \quad 0.0350 \quad 0.004 \quad \ldots\ \right] \\ &+\left[\ 0 \quad 2.0682 \quad 1.0053 \quad 0.2204 \quad 0.0272 \quad \ldots\ \right]. \end{aligned}$$

Imposing that the initial conditions constraint (37) is satisfied yields $\alpha = 8.3853$; the corresponding solution has an error with

$\mathcal{L}_2$-norm 0.0533. The graph of the function y_3 corresponding to such approximation of the input is plotted in Fig. 1. □

Remark 4 (*Data-Driven Free Response Simulation*)**.** A special case of data-driven simulation occurs when the input is zero, i.e., the data-driven simulation of a free response given initial conditions, as formalized in the following problem.

Problem: *Given* $q \in \mathbb{N}$; $d_i \in \mathbb{R}^p$, $i = 0, \dots, q-1$; *and* $t_0 \in \mathbb{I}$ *compute, if it exists, a trajectory* $y' : \mathbb{I} \to \mathbb{R}^p$ *such that* $\text{col}(0, y') \in \mathfrak{B}$ *and*

$$y'^{(i)}(t_0) = d_i \ , \ i = 0, \dots, q-1. \tag{38}$$

The following result is proved analogously to Proposition 4.

Proposition 5. *Denote by* $V_y((i-1)p+1 : ip, :)$ *the ith p-dimensional block-row of* V_y, $i = 1, \dots, L$. *Let K be a basis matrix for* $\ker V_u$. *The free response simulation problem is solvable if and only if there exists* $F \in \mathbb{R}^{n \times \infty}$ *such that*

$$V_y((i-1)p+1 : ip, :)KF\mathcal{D} = V_y(ip+1 : (i+1)p, :)KF \ , \tag{39}$$

$i = 1, \dots, L-1$, *and*

$$V_y KF\mathcal{D}^i \mathfrak{C}(t_0) = d_i \ , \ i = 0, \dots, q-1. \tag{40}$$

If F exists, then $\widetilde{y}' = V_y(1 : p, :)KF$ *is the Chebyshev representation of* y'. □

7. Comments

Comment 1 (*Other Approaches to Continuous-Time Data-Driven Control*)**.** The result closest to ours in problem formulation is in [18]. A fundamental difference is that in [18] at least one *state* trajectory is measured (see section IV.B therein); the state is used also for defining the initial conditions of the simulation (see formula (15) therein). In contrast, we use *only i-o* data, and our approach is applicable also when no insight is available into the internal structure of the system. A second difference is in the computations used for simulation: in [18] a system of *time-varying differential equations* is numerically solved (see formulas (14), (15), (17) therein). In contrast, in (36)–(37) we solve a system of *linear equations*. The state is assumed to be directly measured also in e.g., [19–22]. Often it is assumed that also the *state derivative* is directly measured (see [19–21,23]), even if this is hardly ever possible except for some mechanical systems. In other cases, the state derivative is *numerically estimated* from the values of x at the sampling instants (see [22]), based on assumptions on the intersample behavior (e.g., piecewise constant inputs). In contrast, we work on measurements of the external variables only; the derivatives of the i-o trajectories necessary for simulation do not need to be directly measured; and no assumptions on the intersample behavior are made. □

Comment 2 (*Computation of Chebyshev Coefficients*)**.** Given $N \in \mathbb{N}$, the Chebyshev grid is the set $\{t_i\}_{i=0,\dots,N}$ defined by $t_i := -\cos(i\pi/N)$, $i = 0, \dots, N$. Given $f \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$, an approximation of the first N Chebyshev coefficients $\widetilde{f}_k = \int_{\mathbb{I}} f(t)C_k(t)w(t)dt$ can be computed directly from samples $f(t_i)$, $i = 0, \dots, N$ by *interpolation* rather than integration, as follows. There exists a unique polynomial interpolant $p = \sum_{i=0}^{N} \widetilde{c}_i C_i$ such that $p(t_i) = f(t_i)$, $i = 0, \dots, N$. For a fixed N, the approximation of f given by *truncation* of its Chebyshev series $\sum_{i=0}^{N} \widetilde{f}_i C_i$ and that given by its Chebyshev interpolant $p = \sum_{i=0}^{N} \widetilde{c}_i C_i$ are close, and $\widetilde{c}_i \to \widetilde{f}_i$ as $N \to \infty$, see Th. 4.2 of [10]. How close the approximation is depends on the smoothness of the function f; for differentiable and analytic functions, the approximations in the ∞-norm are within a factor of 2 of each other, see respectively Th. 7.2 and Th. 8.2 of [10]. If the sampling instants $\{t_i\}_{i=0,\dots,N}$ are not arranged in a Chebyshev grid (e.g., when they are equi-spaced), a least-squares procedure can be used to compute approximations of the Chebyshev interpolant [24]. □

Comment 3 (*Projection Error*)**.** The Chebyshev series of a differentiable function converges "algebraically", see Theorem 7.2 p. 53 of [10]; for analytic functions, "geometrically", see Theorem 8.2 p. 57 of [10]; for $\mathcal{C}^\infty$-functions, the approximation error goes to zero faster than $O(N^{-k})$ for every finite k ("exponential convergence"), see p. 47 of [4]. Similar results can be established for less regular functions defined in Sobolev spaces (see Appendix A.11 of [4] and section 5.5.2 of [4]). Consequently, for a fixed value of N, the accuracy achievable using a truncated Chebyshev series computed from N samples of a function $f \in \mathcal{L}_2(\mathbb{I}, \mathbb{R})$ depends on the smoothness of f. Upper bounds on the approximation error are functions of the maximum value of the function in a bounded subset of $\mathbb{C}$ for analytic functions (see (8.1) of [10]); and of the 1-norm of $\frac{df}{dt}$ for once-differentiable functions (see (7.4) in [10]). Error analysis for functions differentiable more than once can be established using Sobolev spaces techniques, see section 5.5.2 of [4]. □

Comment 4 (*On the Choice of Truncation Index*)**.** We assumed that the full Chebyshev representations of u, y are known ($\mathcal{W}_L$ in (20) has an infinite number of columns). Moreover, we computed the full Chebyshev representation of the simulated trajectories (G in Theorem 4 and Proposition 3, and F in Propositions 3 and 4, have an infinite number of columns). In practice, however, only a finite number of coefficients of the simulated signals can be computed.

In principle, the approximation bounds discussed in Comment 3 can be used to estimate the number of Chebyshev coefficients necessary to achieve a given accuracy. In practice, the following procedure is often used (see pp. 18–20 in [10,13]; and Chapter 3 of [25]): the first N Chebyshev coefficients of f are computed. If several consecutive coefficients up to the Nth fall below machine precision, the accuracy is deemed to be sufficient. If not, N is doubled and the process is repeated. Once a sufficiently large N is found, downsampling can be used to reduce the number of coefficients. Other truncation criteria are illustrated in sections 3.1–3.7 of [25].

Such considerations play an important role in computing a finite submatrix of $\mathcal{W}_L$ whose image equals im $\mathcal{W}_L$ (see Theorem 3). To that purpose, one needs to compute also reasonable approximations of the Chebyshev representations of the derivatives of u and y up to the order L (see (20)). Consequently, the number N of coefficients in the Chebyshev representation of u and y must be such that at least the last $N - L$ coefficients $\widetilde{u}_k$ and $\widetilde{y}_k$ are below machine precision. Then also the last coefficients of the Chebyshev representation of $u^{(i)}$ and $y^{(i)}$, $i = 0, \dots, L-1$ are below machine precision. □

Comment 5 (*Approximation Error for Derivatives*)**.** Unless f is a polynomial of degree smaller than or equal to N, the computation (11) of the Chebyshev representation of the derivative incurs an approximation error. Indeed, differentiation and projection do not commute (see section 2.4.2 of [4]), and the error $\frac{d}{dt}(\pi_N(f)) - \pi_N\left(\frac{d}{dt}f\right)$ is nonzero. An upper bound on its norm when f and $f^{(1)}$ are absolutely continuous is given in the following result, whose proof (omitted due to space limitations) uses Theorem 7.2 p. 53 of [10].

Proposition 6. *Let $f \in \mathcal{L}_2(\mathbb{I},\mathbb{R})$ and let $N \in \mathbb{N}$, $N \geq 1$. Assume that f and $f^{(1)}$ are absolutely continuous, and that $\left\| \frac{d^2 f}{dt^2} \right\|_1 = \int_{-1}^{1} \left| \frac{d^2 f}{dt^2}(\tau) \right| d\tau$ is finite. Then the 2-norm of the error is less than or equal to $\frac{2\left\| \frac{d^2 f}{dt^2} \right\|_1}{\sqrt{\pi(N-1)}}$.*

This result can be generalized to the case when f is differentiable more than once, using Sobolev spaces of $\mathcal{L}_2$-functions with a certain number of derivatives in the sense of distributions, see Lemma 2.3 p. 75 in [26]. Proposition 6 can also be used to analyze the situation when the number N of Chebyshev coefficients of the data is fixed, and resampling is not an option. □

Comment 6 (*Chebyshev Representations from Noisy Data*)**.** In practice the problem arises of computing a Chebyshev representation for a function f from samples affected by noise, i.e., from $\widehat{f}(t_i) = f(t_i) + \epsilon_i$ rather than $f(t_i)$, $i = 0, \ldots, N$. Recently [24] two least-squares algorithms have been devised to compute a polynomial approximant from noisy data. Error bounds for the ∞-norm error of the approximant, proportional to the noise variance and inversely proportional to the square of the number of samples, are available. □

Comment 7 (*Accuracy of Simulation*)**.** The number of coefficients necessary to accurately represent a particular input function varies according to how "complex" it is. Moreover, a good Chebyshev approximation of the corresponding output usually requires at least as many non-negligible coefficients as the input function; for example, if the input is a linear combination of exponentials, generically also all natural frequencies of the system will be present in the output. Consequently, in simulations N must be adjusted *dynamically*. Its initial value depends on the complexity of the input function. Then Eqs. (36)–(37) are solved, and the last few Chebyshev coefficients of the corresponding solution y' are checked. If they fall below machine precision, then an accurate representation of the output has been obtained. Otherwise, N is doubled (see Comment 4) and a solution of (36)–(37) for the new value of N is computed. The process continues until an acceptable approximation of y' has been computed. This process is analogous to the "lazy evaluation" used in spectral methods for the solution of differential equations, see section 3 of [13]. □

8. Conclusions

The results illustrated in this paper complement the definition and the characterization of persistency of excitation provided in [5]. We used Chebyshev representations and their properties to *represent all* possible smooth system trajectories in terms of a sufficiently informative one (Theorem 3). We have also devised procedures to *compute* the Chebyshev representation of any smooth system trajectory from that of a sufficiently informative one (Theorems 4 and 5). We used our approach to solve two problems in data-driven simulation, namely the computation of free and forced responses given the initial conditions.

CRediT authorship contribution statement

P. Rapisarda: Conceptualization, Methodology, Writing. **M.K. Çamlibel:** Conceptualization, Methodology, Writing. **H.J. van Waarde:** Conceptualization, Methodology, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] J.C. Willems, P. Rapisarda, I. Markovsky, B.L.M. De Moor, A note on persistency of excitation, Systems Control Lett. 54 (4) (2005) 325–329.

[2] I. Markovsky, P. Rapisarda, Data-driven simulation and control, Internat. J. Control 81 (12) (2008) 1946–1959.

[3] I. Markovsky, J.C. Willems, P. Rapisarda, B.L.D. Moor, Data driven simulation with applications to system identification, IFAC Proc. Vol. 38 (1) (2005) 970–975.

[4] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang, Spectral Methods: Fundamentals in Single Domains, Scientific Computation, Springer-Verlag, Berlin Heidelberg, 2006.

[5] P. Rapisarda, M.K. Çamlibel, H.J. van Waarde, A persistency of excitation condition for continuous-time systems, IEEE Control Syst. Lett. 7 (2023) 589–594.

[6] T. Mühlpfordt, R. Findeisen, V. Hagenmeyer, T. Faulwasser, Comments on truncation errors for polynomial chaos expansions, IEEE Control Syst. Lett. 2 (1) (2018) 169–174.

[7] G. Pan, R. Ou, T. Faulwasser, On a stochastic fundamental lemma and its use for data-driven optimal control, IEEE Trans. Automat. Control (2022) 1–16.

[8] J. Polderman, J. Willems, Introduction to mathematical systems theory: a behavioral approach, in: Texts in Applied Mathematics, Vol. 26, Springer-Verlag, Berlin, 1997.

[9] P. Rapisarda, J. Willems, State maps for linear systems, SIAM J. Control Optim. 35 (3) (1997) 1053–1091.

[10] L.N. Trefethen, Approximation Theory and Approximation Practice, Extended Edition, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019.

[11] J.P. Boyd, Chebyshev and Fourier Spectral Methods, Dover, 2001.

[12] G. Sansone, Orthogonal functions, in: Vol. 9 of Pure and Applied Mathematics, Interscience Publishers, New York-London, 1959.

[13] J.L. Aurentz, L.N. Trefethen, Chopping a Chebyshev series, ACM Trans. Math. Software 43 (4) (2017).

[14] G. Tao, Adaptive Control Design and Analysis, John Wiley & Sons Ltd, 2003.

[15] A. Padoan, G. Scarciotti, A. Astolfi, A geometric characterization of the persistence of excitation condition for the solutions of autonomous systems, IEEE Trans. Automat. Control 62 (11) (2017) 5666–5677.

[16] J.C. Willems, From time series to linear system, part I. Finite dimensional linear time invariant systems, Automatica 22 (5) (1986) 561–580.

[17] T. Driscoll, N. Hale, L. Trefethen, Chebfun Guide, Pafnuty Publications, Oxford, 2014.

[18] V.G. Lopez, M.A. Müller, On a continuous-time version of Willems' lemma, in: 2022 IEEE 61st Conference on Decision and Control, CDC, 2022, pp. 2759–2764.

[19] A. Bisoffi, C.D. Persis, P. Tesi, Data-driven control via Petersen's lemma, Automatica 145 (2022) 110537.

[20] R. Strässer, J. Berberich, F. Allgöwer, Data-driven control of nonlinear systems: Beyond polynomial dynamics, in: 60th IEEE Conference on Decision and Control, CDC, 2021, pp. 4344–4351.

[21] C.D. Persis, P. Tesi, Formulas for data-driven control: Stabilization, optimality, and robustness, IEEE Trans. Automat. Control 65 (3) (2020) 909–924.

[22] J. Berberich, S. Wildhagen, M. Hertneck, F. Allgöwer, Data-driven analysis and control of continuous-time systems under aperiodic sampling, IFAC-PapersOnLine 54 (2021) 210–215, (Extended version at arXiv:2011.09221).

[23] T. Dai, M. Sznaier, A semi-algebraic optimization approach to data-driven control of continuous-time nonlinear systems, IEEE Control Syst. Lett. 5 (2) (2021) 487–492.

[24] Y. Natsukasa, 2023, personal communication.

[25] J.P. Boyd, Solving Transcendental Equations, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[26] C. Canuto, A. Quarteroni, Approximation results for orthogonal polynomials in Sobolev spaces, Math. Comp. 38 (157) (1982) 67–86.